arXiv:1109.6493v1 [math.ST] 29 Sep 2011

Improved estimation in a non-Gaussian parametric regression

Pchelintsev Evgeny*

September 30, 2011

Abstract

The paper considers the problem of estimating the parameters in a continuous time regression model with a non-Gaussian noise of pulse type. The noise is specified by the Ornstein–Uhlenbeck process driven by the mixture of a Brownian motion and a compound Poisson process. Improved estimates for the unknown regression parameters, based on a special modification of the James–Stein procedure with smaller quadratic risk than the usual least squares estimates, are proposed. The developed estimation scheme is applied for the improved parameter estimation in the discrete time regression with the autoregressive noise depending on unknown nuisance parameters.

Keywords Non-Gaussian parametric regression · Improved estimates · Pulse noise · Ornstein–Uhlenbeck process · Quadratic risk · Autoregressive noise

Mathematics Subject Classification (2010) 62H12 - 62M10

*Department of Mathematics and Mechanics, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia, and Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS, Université de Rouen, Avenue de l'Université BP.12, 76800 Saint Etienne du Rouvray Cedex, France, e-mail: evgen-pch@yandex.ru

1 Introduction

In 1961, James and Stein proposed shrinkage estimates which outperform in mean square accuracy the maximum likelihood estimates in the problem of estimating the mean of a multidimensional Gaussian vector with unit covariance matrix (James and Stein 1961). This result stimulated the development of the theory of improved estimation for different regression models with dependent errors. Fourdrinier and Strawderman and Fourdrinier and Wells solved the problem of improved parametric estimation in the regression with dependent non-Gaussian observations under spherically symmetric distributions of the noise (Fourdrinier and Strawderman 1996; Fourdrinier and Wells 1994). Fourdrinier and Pergamenshchikov and Konev and Pergamenchtchikov investigated the problem of improved estimation in nonparametric setting (Fourdrinier and Pergamenshchikov 2007; Konev and Pergamenchtchikov 2010).

In this paper, we consider the problem of improved parametric estimation for a continuous time regression with dependent non-Gaussian noise of pulse type. The noise is specified by the Ornstein–Uhlenbeck process which is known to capture important distributional deviation from Gaussianity and to be appropriate for modelling various dependence structures (Barndorff-Nielsen and Shephard 2001).

Consider a regression model satisfying the equation

$$dy_t = \sum_{j=1}^{p} \theta_j \phi_j(t) dt + d\xi_t, \quad 0 \le t \le n, \tag{1}$$

where $\theta = (\theta_1, ..., \theta_p)'$ (the notation $'$ holds for transposition) is the vector of unknown parameters from a compact set $\Theta \subset \mathbb{R}^p$; $(\phi_j(t))_{1\le j\le p}$ are one-periodic $[0, +\infty) \to \mathbb{R}$ functions, orthonormal in the space $\mathcal{L}_2[0, 1]$. The noise $(\xi_t)_{t\ge 0}$ in (1) is assumed to be a non-Gaussian Ornstein–Uhlenbeck process obeying the following stochastic differential equation

$$d\xi_t = a\xi_t dt + du_t, \tag{2}$$

where $a \le 0$, $(u_t)_{t\ge 0}$ is a Levy process which is the mixture

$$u_t = \varrho_1 w_t + \varrho_2 z_t \tag{3}$$

of a standard Brownian motion $(w_t)_{t\geq 0}$ and a compound Poisson process $(z_t)_{t\geq 0}$ defined as

$$z_t = \sum_{j=1}^{N_t} Y_j, \tag{4}$$

where $(N_t)_{t\geq 0}$ is a Poisson process with the intensity $\lambda > 0$, and $(Y_j)_{j\geq 1}$ is a sequence of i.i.d. Gaussian random variables with parameters (0,1). The noise parameters a, ϱ_1, ϱ_2 and λ are unknown.

The problem is to construct improved estimates for the unknown vector parameter θ on the basis of observations $(y_t)_{0\leq t\leq n}$, which have higher precision as compared with the least squares estimates (LSE).

It will be observed that the regression model (1) is conditionally Gaussian given the σ-algebra $\mathcal{G} = \sigma\{N_t, t \geq 0\}$ generated by the Poisson process. In Section 3 it is shown that the problem of estimating parameter θ in (1) can be reduced to that of estimating the mean in a conditionally Gaussian distribution with a random covariance matrix depending on the unknown nuisance parameters. This enables one to construct a shrinkage estimates for unknown parameters $(\theta_1, \ldots, \theta_p)$ in (1). The main result is given in Theorem 3.1 which claims that the proposed estimate has less risk than the LSE.

In Section 2 we propose a special modification of James–Stein procedure for solving the problem of estimating the mean in a conditionally Gaussian distribution. This procedure allows one to control the mean square accuracy of estimates. It is shown (Theorem 2.1) that this estimate has less mean square risk than the usual LSE.

The rest of the paper is organized as follows. In Section 4 we apply Theorem 2.1 to the problem of parameter estimation in a discrete time regression under a Gaussian autoregressive noise with unknown parameters. Appendix contains some technical results.

2 On improved estimation in a conditionally Gaussian regression

In Section 3, we will shown that the initial problem of estimating the parameters $(\theta_1, \ldots, \theta_p)$ in model (1) reduces to the following one. Suppose that the observation Y is a p-dimensional random vector which obeys the equation

$$Y = \theta + \xi, \tag{5}$$

where θ is an unknown constant vector parameter from some compact set $\Theta \subset \mathbb{R}^p$, ξ is a conditionally Gaussian random vector with zero mean and the covariance matrix $\mathcal{D}(\mathcal{G})$, i.e. $Law(\xi|\mathcal{G}) = \mathcal{N}_p(0, \mathcal{D}(\mathcal{G}))$, where $\mathcal{G}$ is some fixed σ-algebra.

The problem is to estimate θ.

Consider a shrinkage estimate for θ of the form

$$\theta^* = \left(1 - \frac{c}{\|Y\|}\right) Y, \tag{6}$$

where c is a positive constant which will be specified later.

The choice of this estimate (6) is motivated by the need to control the quadratic risk

$$R(\theta, \tilde{\theta}) = \mathbf{E}_\theta \|\theta - \tilde{\theta}\|^2$$

in the case of conditionally Gaussian model (5).

It will be observed that such control can not be provided by the ordinary James–Stein estimate (obtained from (6) by the change of $\|Y\|$ to $\|Y\|^2$).

In order to get an explicit upper bound for the quadratic risk of estimate (6) we impose some conditions on the random covariance matrix $\mathcal{D}(\mathcal{G})$.

Assume that

$(\mathbf{C_1})$ *There exists a positive constant λ_*, such that the minimal eigenvalue of matrix $\mathcal{D}(\mathcal{G})$ satisfies the inequality*

$$\lambda_{min}(\mathcal{D}(\mathcal{G})) \ge \lambda_* \quad \text{a.s.}$$

$(\mathbf{C_2})$ *The maximal eigenvalue of the matrix $\mathcal{D}(\mathcal{G})$ is bounded on some compact set $\Theta \subset \mathbb{R}^p$ from above in the sense that*

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta \lambda_{max}(\mathcal{D}(\mathcal{G})) \le a^*,$$

where a^* is a known positive constant.

Further we will introduce some notation. Let denote the difference of the risks of estimate (6) and LSE $\hat{\theta} = Y$ as

$$\Delta(\theta) := R(\theta^*, \theta) - R(\hat{\theta}, \theta).$$

We will need also the following constant

$$\gamma_p = \frac{\sum_{j=0}^{p-2} 2^{\frac{j-1}{2}} (-1)^{p-j} \mu^{p-1-j} \Gamma\left(\frac{j+1}{2}\right) - (-\mu)^p I(\mu)}{2^{p/2-1} \Gamma\left(\frac{p}{2}\right) d},$$

where $\mu = d/\sqrt{a^*}$,

$$I(a) = \int_0^\infty \frac{\exp(-r^2/2)}{a+r} dr \quad \text{and} \quad d = \sup\{\|\theta\| : \theta \in \Theta\}.$$

Theorem 2.1. *Let the noise ξ in* (5) *have a conditionally Gaussian distribution $\mathcal{N}_p(0, \mathcal{D}(\mathcal{G}))$ and its covariance matrix $\mathcal{D}(\mathcal{G})$ satisfy conditions* $(\mathbf{C_1}), (\mathbf{C_2})$ *with some compact set $\Theta \subset \mathbb{R}^p$. Then the estimator* (6) *with $c = (p-1)\lambda_* \gamma_p$ dominates the LSE $\hat{\theta}_{ML}$ for any $p \geq 2$, i.e.*

$$\sup_{\theta \in \Theta} \Delta(\theta) \leq -[(p-1)\lambda_* \gamma_p]^2.$$

Proof. First we will find the lower bound for the random variable $\|Y\|^{-1}$.

Lemma 2.2. *Under the conditions of Theorem 2.1*

$$\inf_{\theta \in \Theta} \mathbf{E}_\theta \frac{1}{\|Y\|} \geq \gamma_p.$$

The proof of lemma is given in the Appendix.

In order to obtain the upper bound for $\Delta(\theta)$ we will adjust the argument in the proof of Stein's lemma (James and Stein 1961) to the model (5) with a random covariance matrix.

We represent the risks of LSE and of (6) as

$$R(\hat{\theta}, \theta) = \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2 = \mathbf{E}_\theta (\mathbf{E}\|\hat{\theta} - \theta\|^2 | \mathcal{G}) = \mathbf{E}_\theta tr \mathcal{D}(\mathcal{G});$$

$$R(\theta^*, \theta) = R(\hat{\theta}, \theta) + \mathbf{E}_\theta [\mathbf{E}((g(Y) - 1)^2 \|Y\|^2 | \mathcal{G})]$$

$$+ 2\sum_{j=1}^p \mathbf{E}_\theta [\mathbf{E}((g(Y) - 1) Y_j (Y_j - \theta_j) | \mathcal{G})],$$

where $g(Y) = 1 - c/\|Y\|$.

Denoting $f(Y) = (g(Y) - 1)Y_j$ and applying the conditional density of distribution of a vector Y with respect to σ-algebra $\mathcal{G}$

$$p_Y(x|\mathcal{G}) = \frac{1}{(2\pi)^{p/2} \sqrt{\det \mathcal{D}(\mathcal{G})}} \exp\left(-\frac{(x-\theta)' \mathcal{D}^{-1}(\mathcal{G})(x - \theta)}{2} \right),$$

one gets

$$I_j := \mathbf{E}(f(Y)(Y_j - \theta_j)|\mathcal{G}) = \int_{\mathbb{R}^p} f(x)(x - \theta_j) p_Y(x|\mathcal{G}) dx, \quad j = \overline{1, p}.$$

Making the change of variable $u = \mathcal{D}^{-1/2}(\mathcal{G})(x-\theta)$ and assuming $\tilde{f}(u) = f(\mathcal{D}^{1/2}(\mathcal{G})u+\theta)$, one finds that

$$I_j = \frac{1}{(2\pi)^{p/2}} \sum_{l=1}^{p} \langle \mathcal{D}^{1/2}(\mathcal{G}) \rangle_{jl} \int_{\mathbb{R}^p} \tilde{f}(u) u_l \exp\left(-\frac{\|u\|^2}{2}\right) du, \quad j = \overline{1,p},$$

where $\langle A \rangle_{ij}$ denotes the (i,j)-th element of matrix A. These quantities can be written as

$$I_j = \sum_{l=1}^{p}\sum_{k=1}^{p} \mathbf{E}(< \mathcal{D}^{1/2}(\mathcal{G}) >_{jl} < \mathcal{D}^{1/2}(\mathcal{G}) >_{kl} \frac{\partial f}{\partial u_k}(u)|_{u=Y}|\mathcal{G}), \quad j = \overline{1,p}.$$

Thus, the risk for an estimator (6) takes the form

$$R(\theta^*,\theta) = R(\hat{\theta},\theta) + \mathbf{E}_\theta((g(Y)-1)^2\|Y\|^2)$$
$$+2\mathbf{E}_\theta\left(\sum_{j=1}^{p}\sum_{l=1}^{p}\sum_{k=1}^{p} < \mathcal{D}^{1/2}(\mathcal{G}) >_{jl} < \mathcal{D}^{1/2}(\mathcal{G}) >_{kl} \frac{\partial}{\partial u_k}[(g(u)-1)u_j]|_{u=Y}\right).$$

Therefore, one has

$$R(\theta^*,\theta) = R(\hat{\theta},\theta) + \mathbf{E}_\theta W(Y),$$

where

$$W(z) = c^2 + 2c\frac{z'\mathcal{D}(\mathcal{G})z}{\|z\|^3} - 2tr\mathcal{D}(G)c\frac{1}{\|z\|}.$$

This implies that

$$\Delta(\theta) = \mathbf{E}_\theta W(Y).$$

Since $z'Az \le \lambda_{max}(A)\|z\|^2$, one comes to the inequality

$$\Delta(\theta) \le c^2 - 2c\mathbf{E}_\theta \frac{tr\mathcal{D}(\mathcal{G}) - \lambda_{max}(\mathcal{D}(\mathcal{G}))}{\|Y\|}.$$

From here, it follows that

$$\Delta(\theta) \le c^2 - 2c\sum_{i=2}^{p} \mathbf{E}_\theta \frac{\lambda_i(\mathcal{D}(\mathcal{G}))}{\|Y\|}.$$

Taking into account the condition $(\mathbf{C_1})$ and the Lemma 2.2, one obtains

$$\Delta(\theta) \le c^2 - 2(p-1)\lambda_*\gamma_p c =: \phi(c).$$

Minimizing the function $\phi(c)$ with respect to c, we come to the desired result, i.e.

$$\Delta(\theta) \leq -[(p-1)\lambda_* \gamma_p]^2.$$

Hence Theorem 2.1.

$\square$

Corollary 2.3. *Let in* (5) *the noise* $\xi \sim \mathcal{N}_p(0, D)$ *with the positive definite non random covariance matrix* $D > 0$ *and* $\lambda_{min}(D) \geq \lambda_* > 0$*. Then the estimator* (6) *with* $c = (p-1)\lambda_* \gamma_p$ *dominates the LSE for any* $p \geq 2$ *and compact set* $\Theta \subset \mathbb{R}^p$*, i.e.*

$$\sup_{\theta \in \Theta} \Delta(\theta) \leq -[(p-1)\lambda_* \gamma_p]^2.$$

Remark 2.1. *Note that if* $D = \sigma^2 I_p$ *then*

$$\sup_{\theta \in \Theta} \Delta(\theta) \leq -[(p-1)\sigma^2 \gamma_p]^2.$$

Corollary 2.4. *If* $\xi \sim \mathcal{N}_p(0, I_p)$ *and* $\theta = 0$ *in model* (5) *then the risk of estimate* (6) *is given by the formula*

$$R(0, \theta^*) = p - \left[\frac{(p-1)\Gamma((p-1)/2)}{\sqrt{2}\Gamma(p/2)} \right]^2 =: r_p. \tag{7}$$

By applying the Stirling's formula for the Gamma function

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} \exp(-x) \, (1 + o(1))$$

one can check that $r_p \to 0.5$ as $p \to \infty$. The behavior of the risk (7) for small values of p is shown in Fig.1. It will be observed that in this case the risk of the James–Stein estimate $\hat{\theta}_{JS}$ remains constant for all $p \geq 3$, i.e.

$$R(0, \hat{\theta}_{JS}) = 2$$

and the risk of the LSE $\hat{\theta}$ is equal to p and tends to infinity as $p \to \infty$.

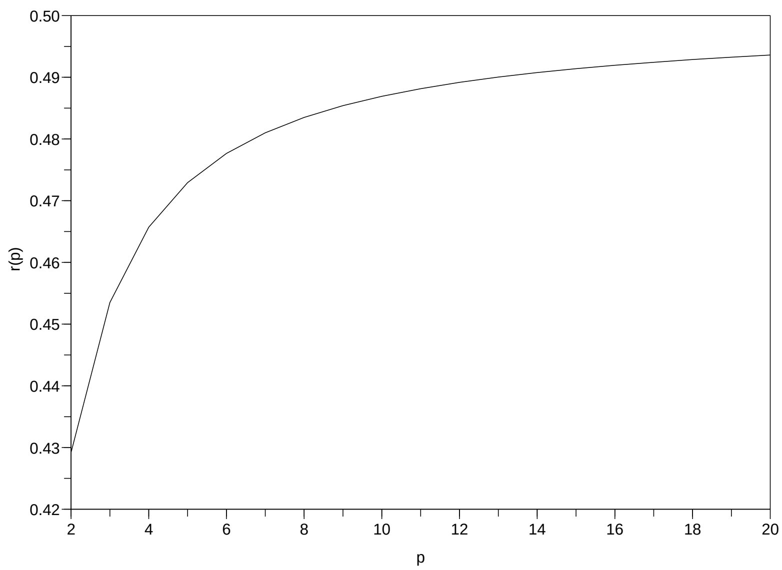

Figure 1: Risk of θ^* at $\theta = 0$.

3 Improved estimation in a non-Gaussian Ornstein–Uhlenbeck–Levy regression model

In this section we use the estimate (6) to a non-Gaussian continuous time regression model to construct an improved estimate of the unknown vector parameter θ. To this end we reduce first the initial continuous time regression model (1) to a discrete time model of the form (5) with a conditionally Gaussian noise.

A commonly used estimator of an unknown vector θ in model (1) on the basis of observations $(y_t)_{0\le t\le n}$ is the LSE $\hat\theta = (\hat\theta_1, \dots, \hat\theta_p)'$ with the components

$$\hat\theta_j = \frac{1}{n}\int_0^n \phi_j(t)dy_t, \quad j = \overline{1,p}.$$

From here and (1), one has

$$\hat\theta = \theta + n^{-1/2}\zeta(n), \tag{8}$$

where $\zeta(n)$ is the random vector with the coordinates

$$\zeta_j(n) = n^{-1/2} \int_0^n \phi_j(t) d\xi_t.$$

Note that the vector $\zeta(n)$ has a conditionally Gaussian distribution with a zero mean and conditional covariance matrix $V_n(\mathcal{G}) = cov(\zeta(n), \zeta(n)'|\mathcal{G})$ with the elements

$$v_{ij}(n) = \mathbf{E}(\zeta_i(n)\zeta_j(n) \mid \mathcal{G}).$$

Thus the initial problem of estimating parameter θ in (1) can be reduced to the that of estimating parameter θ in a conditionally Gaussian regression model (8).

Theorem 3.1. *Let the regression model be given by the equations* (1)–(4), $\varrho_1 > 0$. *Then, for any $n \geq 1$ and $p \geq 2$, the estimator of θ*

$$\theta^* = \left(1 - \frac{\varrho_1^2 (p-1)\gamma_p}{n \|\hat{\theta}\|}\right) \hat{\theta},$$

dominates the LSE $\hat{\theta}$:

$$\sup_{\theta \in \Theta} \Delta(\theta) \leq - \left[\frac{\varrho_1^2 (p-1)\gamma_p}{n}\right]^2.$$

To prove this theorem one can apply Theorem 2.1. To this end it suffices to check conditions $(\mathbf{C_1})$, $(\mathbf{C_2})$ on the matrix $V_n(\mathcal{G})$. The verification of conditions $(\mathbf{C_1})$ and $(\mathbf{C_2})$ is given in the Appendix.

4 Improved estimation in an autoregression

In this section we consider the problem of improved estimating the unknown mean of a multivariate normal distribution when the dispersion matrix is unknown and depends on some nuisance parameters. The models of autoregressive type are widely used in time series analysis (Anderson 1994; Brockwell and Davis 1991).

Let the noise $\xi = (\xi_1, \ldots, \xi_p)'$ in (5) be described by a Gaussian autoregression process

$$\xi_k = a\xi_{k-1} + \varepsilon_k, \ k = \overline{1,p}, \tag{9}$$

where $|a| < 1$, $\mathbf{E}\xi_0 = 0$ and $\varepsilon_1, \ldots, \varepsilon_p$ are independent Gaussian (0,1) random variables. Assume that the parameter a in (9) is unknown and belongs to interval $[-\alpha, \alpha]$, where $0 < \alpha < 1$ is known number.

It is easy to check that the covariance of the noise ξ has the form

$$D(a) = \frac{1}{1-a^2} \begin{pmatrix} 1 & a & \ldots & a^{p-1} \\ a & 1 & \ldots & a^{p-2} \\ & \ddots & & \\ a^{p-1} & a^{p-2} & \ldots & 1 \end{pmatrix}$$

Proposition 4.1. *Let the noise ξ in (5) be specified by equation (9) with $a \in [-\alpha, \alpha]$. Then, for any $p > 1/(1-\alpha)^2$, the LSE is dominated by the estimate*

$$\theta^* = \left(1 - \left(p - \frac{1}{(1-\alpha)^2}\right) \frac{\gamma_p}{\|Y\|}\right) Y$$

in the sense that

$$\sup_{\theta \in \Theta} \Delta(\theta) \leq - \left(p - \frac{1}{(1-\alpha)^2}\right)^2 \gamma_p^2.$$

Proof. We note that $trD(a) = p/(1-a^2)$. Now we will estimate of the maximal eigenvalue of matrix $D(a)$. From the definition

$$\lambda_{max}(D(a)) = \sup_{\|z\|=1} z' D(a) z$$

one has

$$z'D(a)z = \sum_{i=1}^{p} \sum_{j=1}^{p} < D(a) >_{ij} z_i z_j = \frac{1}{1-a^2} \left(1 + 2 \sum_{i=1}^{p-1} \sum_{j=1}^{p-i} a^j z_i z_{j+i}\right)$$

$$= \frac{1}{1-a^2} \left(1 + 2 \sum_{j=1}^{p-1} a^j \sum_{i=1}^{p-j} z_j z_{i+j}\right).$$

Applying the Cauchy–Bunyakovskii inequality yields

$$\lambda_{max}(D(a)) \leq \frac{1}{1-\alpha^2} \left(1 + 2 \sum_{j=1}^{\infty} \alpha^j\right) = \frac{1}{(1-\alpha)^2}.$$

Thus,

$$trD(a) - \lambda_{max}(D(a)) \geq p - \frac{1}{(1-\alpha)^2}.$$

By applying Theorem 2.1 we come to the assertion of Proposition 4.1.

$\square$

5 Appendix

5.1 Proof of the Lemma 2.2.

Proof. From (5), one has

$$J = \mathbf{E}_\theta \frac{1}{\|Y\|} = \mathbf{E}_\theta \frac{1}{\|\theta + \xi\|} \geq \mathbf{E}_\theta \frac{1}{d + \|\xi\|}.$$

Taking the repeated conditional expectation and noting that the random vector ξ is conditionally Gaussian with zero mean, one gets

$$J \geq \mathbf{E}_\theta \frac{1}{(2\pi)^{p/2}\sqrt{det\mathcal{D}(\mathcal{G})}} \int_{\mathbb{R}^p} \frac{\exp(-x'\mathcal{D}(\mathcal{G})^{-1}x/2)}{d + \|x\|} dx.$$

Making the change of variable $u = \mathcal{D}(\mathcal{G})^{-1/2}x$ and applying the estimation $u'\mathcal{D}(\mathcal{G})u \leq \lambda_{max}(\mathcal{D}(\mathcal{G}))\|u\|^2$, we find

$$J \geq \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} \frac{\exp(-\|u\|^2/2)}{d + \sqrt{\lambda_{max}(\mathcal{D}(\mathcal{G}))}\|u\|} du.$$

Further making the spherical changes of the variables yields

$$J \geq \frac{1}{2^{p/2-1}\Gamma(p/2)} \mathbf{E}_\theta \int_0^\infty \frac{r^{p-1}\exp(-r^2/2)}{d + \sqrt{\lambda_{max}(\mathcal{D}(\mathcal{G}))}r} dr.$$

From here by applying the Jensen and Cauchy–Bunyakovskii inequalities and by the condition $(\mathbf{C_2})$, we obtain

$$J \geq \frac{\mu}{2^{p/2-1}\Gamma(p/2)d} \int_0^\infty \frac{r^{p-1}\exp(-r^2/2)}{\mu + r} dr = \gamma_p.$$

This leads to the assertion of Lemma 2.2.

$\square$

5.2 The verification of the conditions $(\mathbf{C_1})$ and $(\mathbf{C_2})$ on the matrix $V_n(\mathcal{G})$.

Now we establish some properties of a stochastic integral

$$I_t(f) = \int_0^t f_s d\xi_s, \quad 0 \le t \le n$$

with respect to the process (2). We will need some notations. Let us denote

$$\varepsilon_f(t) = a \int_0^t \exp\{a(t-v)\}\, f(v)\,(1 + \exp(2av))\, dv\,,$$

where f is $[0,+\infty) \to \mathbb{R}$ function integrated on any finite interval. We introduce also the following transformation

$$\tau_{f,g}(t) = \frac{1}{2}\int_0^t \Big(2f(s)g(s) + \varepsilon^*_{f,g}(s)\Big)\, ds$$

of square integrable $[0,+\infty) \to \mathbb{R}$ functions f and g. Here

$$\varepsilon^*_{f,g}(t) = f(t)\varepsilon_g(t) + \varepsilon_f(t)g(t)\,.$$

Proposition 5.1. *If f and g are functions from $\mathcal{L}_2[0,n]$ then*

$$\mathbf{E}\, I_t(f)I_t(g) = \varrho^*\, \tau_{f,g}(t) \tag{10}$$

where $\varrho^* = \varrho_1^2 + \lambda\varrho_2^2$.

Proof. Noting that the process $I_t(f)$ satisfies the stochastic equation

$$dI_t(f) = af(t)\xi_t dt + f(t)du_t\,, \quad I_0(f) = 0\,,$$

and applying the Ito formula, one obtains (10). Hence Proposition 5.1.

□

Corollary 5.2. *If f is function from $\mathcal{L}_2[0,n]$ then*

$$\mathbf{E}\, I_n^2(f) \le 3\varrho^* \int_0^n f^2(t)dt. \tag{11}$$

Further, for an integrated $[0,+\infty) \to \mathbb{R}$ function f, we define the function

$$L_f(x,z) = a\exp(ax)\,\left(f(z) + a\int_0^x \exp(av)\,f(v+z)\,dv\right).$$

Let $(T_l)_{l\geq 1}$ be the jump times of the Poisson process $(N_t)_{t\geq 0}$, i.e.

$$T_l = \inf\{t \geq 0 : N_t = l\}.$$

Proposition 5.3. *Let f and g be bounded left-continuous $[0,\infty)\times\Omega \to \mathbb{R}$ functions measurable with respect to $\mathcal{B}[0,+\infty)\bigotimes\mathcal{G}$ (the product σ algebra created by $\mathcal{B}[0,+\infty)$ and $\mathcal{G}$). Then*

$$\mathbf{E}\,(I_t(f)|\mathcal{G}) = 0$$

and

$$\mathbf{E}\,(I_t(f)\,I_t(g)|\mathcal{G}) = \varrho_1^2\tau_{f,g}(t) + \varrho_2^2\sum_{l\geq 1} f(T_l)g(T_l)\mathbf{1}_{(T_l\leq t)}$$
$$+\,\varrho_2^2\sum_{l\geq 1}\int_0^t \left(f(s)L_g(s-T_l,T_l) + g(s)L_f(s-T_l,T_l)\right)\mathbf{1}_{(T_l\leq s)}ds.$$

Proof. By the Ito formula one has

$$I_t(f)\,I_t(g) = \int_0^t (\varrho_1^2 f(s)g(s) + a(f(s)I_s(g) + g(s)I_s(f))\xi_s)\mathrm{d}s$$
$$+\,\varrho_2^2\sum_{l\geq 1} f(T_l)\,g(T_l)\,Y_l^2\mathbf{1}_{\{T_l\leq t\}}$$
$$+\int_0^t (f(s)I_{s-}(g) + g(s)I_{s-}(f))\mathrm{d}u_s\,.$$

Taking the conditional expectation $\mathbf{E}\,(\cdot|\mathcal{G})$, on the set $\{T_l > t\}$, yields

$$\mathbf{E}\,(I_t(f)\,I_t(g)|\mathcal{G}) = \int_0^t \varrho_1^2 f(s)g(s)\mathrm{d}s + \varrho_2^2\sum_{l\geq 1} f(T_l)\,g(T_l)\;\mathbf{1}_{\{T_l\leq t\}}$$
$$+\,a\int_0^t \left(f(s)\mathbf{E}(I_s(g)\xi_s|\mathcal{G}) + g(s)\mathbf{E}(I_s(f)\xi_s|\mathcal{G})\right)\mathrm{d}s\,.$$

It is easy to check that

$$a\mathbf{E}(I_t(f)\xi_t|\mathcal{G}) = \frac{\varrho_1^2}{2}\varepsilon_f(t) + \varrho_2^2 \sum_{j\geq 1} L_f(t-T_j, T_j)\,\mathbf{1}_{\{T_j\leq t\}}.$$

From here one comes to the desired equality. Hence Proposition 5.3.

$\square$

Thus, in view of $\zeta_j(n) = n^{-1/2}I_n(\phi_j)$ and Proposition 5.3 the elements of covariance matrix $V_n(\mathcal{G})$ can be written as

$$\begin{aligned} v_{ij}(n) &= \frac{\varrho_1^2}{n}\int_0^n \phi_i(t)\phi_j(t)dt \\ &\quad + \frac{\varrho_1^2}{2n}\int_0^n \left(\phi_i(t)\varepsilon_{\phi_j}(t) + \phi_j(t)\varepsilon_{\phi_i}(t)\right)dt + \frac{\varrho_2^2}{n}\sum_{l\geq 1}\phi_i(T_l)\phi_j(T_l)\mathbf{1}_{(T_l\leq n)} \\ &\quad + \frac{\varrho_2^2}{n}\sum_{l\geq 1}\int_0^n \left(\phi_i(t)L_{\phi_j}(t-T_l,T_l) + \phi_j(t)L_{\phi_i}(t-T_l,T_l)\right)\mathbf{1}_{(T_l\leq t)}dt. \end{aligned} \tag{12}$$

Lemma 5.4. *Let $(\xi_t)_{t\geq 0}$ be defined by (2) with $a\leq 0$. Then a matrix $V_n(\mathcal{G}) = (v_{ij}(\phi))_{1\leq i,j\leq p}$ with elements defined by (12), satisfy the following inequality a.s.*

$$\inf_{n\geq 1}\inf_{\|z\|=1} z'V_n(\mathcal{G})z \geq \varrho_1^2.$$

Proof. Notice that by (12) one can the matrix $V_n(\mathcal{G})$ present as

$$V_n(\mathcal{G}) = \varrho_1^2 I_p + F_n + B_n(\mathcal{G}),$$

where F_n is non random matrix with elements

$$f_{ij}(n) = \frac{\varrho_1^2}{2n}\int_0^n \left(\phi_i(t)\varepsilon_{\phi_j}(t) + \phi_j(t)\varepsilon_{\phi_i}(t)\right)dt$$

and $B_n(\mathcal{G})$ is a random matrix with elements

$$\begin{aligned} b_{ij}(n) = \frac{\varrho_2^2}{n}\sum_{l\geq 1}&[\phi_i(T_l)\phi_j(T_l)\mathbf{1}_{(T_l\leq n)} \\ &+ \int_0^n \left(\phi_i(t)L_{\phi_j}(t-T_l,T_l) + \phi_j(t)L_{\phi_i}(t-T_l,T_l)\right)\mathbf{1}_{(T_l\leq t)}dt]. \end{aligned}$$

This implies that

$$z'V_n(\mathcal{G})z = \varrho_1^2 z'z + z'F_n z + z'B_n(\mathcal{G})z \geq \varrho_1^2 z'z,$$

therefore

$$\inf_{\|z\|=1} z'V_n(\mathcal{G})z \geq \varrho_1^2$$

and we come to the assertion of Lemma 5.4.

□

Lemma 5.5. *Let* $(\xi_t)_{t\geq 0}$ *be defined by* (2) *with* $a \leq 0$. *Then a maximal eigenvalue of the matrix* $V_n(\mathcal{G}) = (v_{ij}(n))_{1\leq i,j\leq p}$ *with elements defined by* (12), *satisfy the following inequality*

$$\sup_{n\geq 1}\sup_{\theta\in\Theta} \mathbf{E}_\theta \lambda_{max}(V_n(\mathcal{G})) \leq 3p\varrho^*.$$

where $\varrho^* = \varrho_1^2 + \lambda\varrho_2^2$.

Proof. We note that

$$\mathbf{E}_\theta \lambda_{max}(V_n(\mathcal{G})) \leq \mathbf{E}_\theta tr(V_n(\mathcal{G})) = \sum_{j=1}^{p} \mathbf{E}_\theta \zeta_j^2(n) = \frac{1}{n}\sum_{j=1}^{p} \mathbf{E}_\theta I_n^2(\phi_j).$$

Applying (11) and $\int_0^n \phi_j^2(t)dt = n$, we obtain the desired inequality.
Hence Lemma 5.5.

□

Thus the matrix $V_n(\mathcal{G})$ is positive definite and satisfies for any compact set $\Theta \subset \mathbb{R}^p$, the conditions $(\mathbf{C_1})$ and $(\mathbf{C_2})$ with $\lambda_* = \varrho_1^2$ and $a^* = 3p\varrho^*$.

References

[1] Anderson T.W. (1994) The Statistical Analysis of Time Series. Wiley, New York, London, Sydney, Toronto

[2] Barndorff-Nielsen O.E., Shephard N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics. J. Royal Stat. Soc. B 63:167-241

[3] Brockwell P.J., Davis R.A. (1991) Time Series: Theory and Methods, 2nd edn. Springer, New York

[4] Fourdrinier D., Pergamenshchikov S. (2007) Improved selection model method for the regression with dependent noise, Ann. of the Inst. of Statist. Math. 59(3):435-464

[5] Fourdrinier D., Strawderman W.E. (1996) A paradox concerning shrinkage estimators: should a known scale parameter be replaced by an estimated value in the shrinkage factor? J. Multivariate Anal. 59(2):109-140

[6] Fourdrinier D., Wells M.T. (1994) Comparaison de procédures de sélection d'un modèle de régression: une approche décisionnelle, Comptes Rendus de l'Académie des Sciences Paris, v.319, série I:865-870

[7] James W., Stein C. (1961) Estimation with quadratic loss, in: Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability, Vol. 1, University of California Press, Berkeley, p.361-380

[8] Konev V., Pergamenchtchikov S. (2010) General model selection of a periodic regression with a Gaussian noise. An. Inst. Stat. Math. 62:1083-1111